



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2014

---

## **Frequency, Form and Function of Cleft Constructions in the Swiss SMS Corpus**

Stark, Elisabeth

DOI: <https://doi.org/10.1515/9783110361872.325>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-105518>

Book Section

Published Version

Originally published at:

Stark, Elisabeth (2014). Frequency, Form and Function of Cleft Constructions in the Swiss SMS Corpus.  
In: de Cesare, Anna-Maria. Frequency, Forms and Functions of Cleft Constructions in Romance and  
Germanic. Berlin/München/Boston: De Gruyter, 325-344.

DOI: <https://doi.org/10.1515/9783110361872.325>

# **Frequency, Forms and Functions of Cleft Constructions in Romance and Germanic**

---

Contrastive, Corpus-based studies

Edited by  
Anna-Maria De Cesare

**DE GRUYTER**  
MOUTON

ISBN 978-3-11-036107-0  
e-ISBN (PDF) 978-3-11-036187-2  
e-ISBN (EPUB) 978-3-11-039474-0  
ISSN 1861-4302

**Library of Congress Cataloging-in-Publication Data**

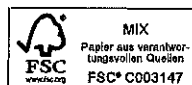
A CIP catalog record for this book has been applied for at the Library of Congress.

**Bibliographic information published by the Deutsche Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;  
detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>.

© 2014 Walter de Gruyter GmbH, Berlin/Munich/Boston  
Typesetting: PTP-Berlin Protago-TEX-Production GmbH, Berlin  
Printing: CPI books GmbH, Leck  
♻️ Printed on acid-free paper  
Printed in Germany

[www.degruyter.com](http://www.degruyter.com)



## Table of Contents

Anna-Maria De Cesare  
Introduction — 1

### Part I. Italian Cleft Constructions in a Contrastive Perspective

Anna-Maria De Cesare  
Cleft constructions in a contrastive perspective. Towards an operational  
taxonomy — 9

Anna-Maria De Cesare, Davide Garassino, Rocío Agar Marco and Laura Baranzini  
Form and frequency of Italian Cleft constructions in a corpus of electronic news.  
A contrastive perspective with French, Spanish, German and English — 49

Davide Garassino  
Cleft sentences. Italian-English in contrast — 101

Laura Baranzini  
Pseudo-cleft sentences. Italian-French in contrast — 139

Rocío Agar Marco  
Pseudo-cleft sentences. Italian-Spanish in contrast — 177

Iørn Korzen  
Cleft sentences. Italian-Danish in contrast — 217

Giovanna Brianti  
Cleft sentences. A translation perspective on Italian and French — 277

### Part II. Romance and Germanic Cleft Constructions in Contrast

Elisabeth Stark  
Frequency, form and function of Cleft constructions in the Swiss SMS  
corpus — 325

Vahram Atayan and Ursula Wienen

Inferential cleft constructions in translation. French *c'est que* in political texts — 345

Volker Gast and Natalia Levshina

Motivating  $w(\mu)$ -Clefts in English and German: A hypothesis-driven parallel corpus study — 377

Subject index — 415

Elisabeth Stark\*

## Frequency, form and function of Cleft constructions in the Swiss SMS corpus

### 1 Introduction

Following Jespersen (1937) and subsequent work, Lambrecht (2001) provides the following definition of a Cleft construction (see also Ferrari et al. 2008: 224–225 for Italian):

A cleft construction [...] is a complex sentence structure consisting of a matrix clause headed by a copula and a relative or relative-like clause whose relativized argument is coindexed with the predicative argument of the copula. Taken together, the matrix and the relative express a logically simple proposition, which can also be expressed in the form of a single clause without a change in truth conditions. (Lambrecht 2001: 466)

We will strictly adhere to this definition and consider Cleft constructions to be only those sentences which can be reduced without any substantial syntactic and semantic change to simple sentences with the same vericonditional properties as the corresponding Cleft sentence; see examples (1) to (4):

- (1) a. It is the wife that decides.  
 b. C' est l' épouse qui décide.  
 c. È la sposa che decide.  
 d. Es ist die Frau, die entscheidet.  
 EXPL 3s.be ART.DEF wife REL 3s.decide

\* I would like to thank Anna-Maria De Cesare for having invited me to her exciting project and workshop, the Swiss National Science Foundation for substantial financial support of the project „SMS communication in Switzerland“, CRSII1\_136230, and the anonymous reviewers for having helped considerably improving the paper. All remaining errors are, of course, mine.

- (2) a. What I want is music.  
 b. Ce que je veux, c' est de la musique.  
 c. Quello che voglio è musica.  
 d. Was ich will, ist Musik.  
 REL 1s 1s.want EXPL 3s.be ART.PART music

(Dufter 2009: 83)

- (3) a. Music is what I want. (Dufter 2009: 89–90)  
 b. Musik ist, was ich will.  
 music 3s.be - REL 1s 1s.want

- (4) That's what is difficult to understand.

The examples in (1) give English, French, Italian and German instances of a so-called *IT-cleft*, which is constituted by an expletive subject (Engl. *it*, Fr. *ce*, G. *es*, also *pro* in pro-drop languages like Standard Italian or Spanish), a tensed form of a copula, the clefted constituent, a *WH*-element and the predicate attributed to the element expressed in the clefted constituent. The examples in (2) feature the same languages and are instances of a so-called *WH-cleft* (or *pseudo-cleft*; we will use the term *WH-cleft* in what follows), beginning with a *WH*-element, the predicate, a tensed form of the copula verb and the clefted constituent. Example (3) shows examples of *reverse WH-clefts*, non-existent in French and Italian according to Lambrecht (2001) or Van den Steen (2005), in which the clefted constituent is fronted before the copula verb, the *WH*-element and the predicate. So-called *THAT-clefts*, shown in (4), are similarly non-existent in French and Italian and begin with a demonstrative or presentative element, referring anaphorically back to the referent of the clefted constituent, before a tensed form of the copula, a *WH*-element and the predicate. For each of these examples, there exist equivalent sentences without the expletive and/or *WH*-elements and the copula: *The wife decides* and its equivalents in the other languages for (1), *I want music* etc. for (2) and (3) and *That is difficult to understand* for (4).

The overall function of clefts is normally associated with textual coherence due to their specific information structure related properties (see below and e.g. Büring 2006: 145; for French *WH-clefts* Krötsch and Sabban 1990; Blanche-Benveniste 2010: 185), and they are no longer considered simple stylistic variants of the corresponding monoclausal propositions (cf. e.g. Engebreten 2012: 125).

In what follows, we will try to give a substantial answer to two research questions:

- 1) Which types of Cleft constructions do we find, how often, and in which language in our SMS data?
- 2) Do we see any language-specific conventionalization tendencies ("possible fossilizations of highly frequent subtypes of clefts", Dufter 2009: 115)?

As far as we know, this study is the first empirical research on the use of Cleft constructions in text messages (SMS). We will take into account clefts in non-dialectal German, Swiss German, French and Italian text messages of the newly established Swiss reference corpus of text messages (cf. [www.sms4science.ch](http://www.sms4science.ch)). Little is known about the use of Cleft constructions in the non-standard varieties of Switzerland, especially Swiss German dialects, and we hope to enlarge our knowledge on this specific type of syntactic structure in a special form of communication, texting (cf. Dürscheid 2002), and of the German dialectal varieties used in Swiss text messages. The paper is structured as follows: after this short introduction, we will provide an overview of the most important findings in the literature concerning the function of Cleft constructions and their language specific distribution in different types of communication forms and/or registers (Section 2). From this, we will formulate three guiding hypotheses for our analysis. Section 3 will briefly introduce our database, the newly established Swiss reference corpus of text messages (cf. [www.sms4science.ch](http://www.sms4science.ch)), before we present our results in Section 4. Section 5 seeks to interpret these findings against the background of preceding studies on clefts in German and Romance languages, and a short conclusion in Section 6 will reiterate the most important insights we have gained from our research.

## 2 Distribution and functions of Cleft constructions: State of the art and hypotheses

Although no research has been conducted into the distribution of Cleft constructions in new forms of communication, especially in computer-mediated communication (CMC),<sup>1</sup> the specialized functional literature gives some insights into their functional and variationist profile, which might help develop some guiding hypotheses about their probable distribution in our data.

<sup>1</sup> A search for "Cleft constructions" and "computer-mediated communication" in the data bases LLBA and MLA returned zero results.

Prince (1978) and Collins (1991) state that IT-clefts occur more often in (formal) writing. They seem to be typically used in written discourse, whereby the more formal conception of the text (type) in question appears more important than the medium (i.e., the phonic or graphical realization; cf. Koch and Oesterreicher [1990] 2011 for this fundamental distinction). This is in line with Roggia's (2008) results on the overall distribution of IT-clefts in French and Italian, based on the phonic corpus C-ORAL-ROM (cf. Roggia 2008: 19). As for the influence of the medium on French Cleft sentence types, contrary to Prince's statement, Dufter (2008: 42) finds that "[...] *c'est*-clefts [= the French equivalent to IT-clefts, ES] occur about 2.5 times more often in speech than in writing [...]" Dufter (2008) based his research on the corpus C-ORAL-ROM for phonic (formal and informal situation types; cf. Cresti and Moneglia 2005) and on FRANTEXT for formal written French. Bearing these divergent statements in mind, we will formulate our *Hypothesis I* as follows: IT-clefts and their equivalents are expected to appear quite rarely in text messages, the latter ones being typical instances of informal communication (cf. the discussion in Stark 2011).

According to Collins (1991), WH-clefts (and reverse WH-clefts) are typically found in spoken language, especially in dialogues between participants who are (close) friends. This means that the rather informal conception of a message favors the occurrence of WH-clefts. Here, we are confronted additionally with a language-specific bias (which holds generally for the use of Cleft sentences, see below): they seem to be quite rare in French (cf. Van den Steen 2005: 277) and German (cf. the results of Gast and Wiechmann 2012 for the use of WH-clefts in English vs. German in the EUROPARL corpus). We thus have to formulate Hypothesis II in two parts: a. WH-clefts should be particularly frequent in text messages; b. WH-clefts are in general rather infrequent in the French, German and probably also Italian subcorpora.

Concerning language-specific preferences for certain Cleft construction (sub-) types, Dufter (2009), in his empirical investigation of translations of English IT-clefts in the EUROPARL-corpus into French, Portuguese, Italian, Spanish and German, notices a decreasing likelihood of translating English IT-clefts with Cleft sentences, in the above-mentioned order. While the equivalents of IT-clefts are used frequently in French and Portuguese, they are used slightly less frequently in Italian and Spanish, and German uses different constructions, especially Verb (V)-Subject (S) word order, to substitute them. (For another empirical study corroborating this view on language specific frequencies and functions of clefts, cf. Engebretsen 2012). In order to explain these empirical findings, Dufter (2009) reminds us of an important functional subcategorization of IT-clefts, arguing against a monofunctional approach (cf. particularly Lambrecht 1994, 2001), which would see focus marking as the only function of IT-clefts. Prince (1978)

distinguishes essentially three sub-types of IT-clefts (see also e.g. Gundel 2008): first, a *Type A*, called *stressed-focus cleft*, where either the clefted constituent constitutes the only narrow focus of the sentence (example [5]) or which may have also multiple foci (example [6]):

- (5) It was Peter who stole the money (not John).  
 (6) It is the money that we have to save and the economy that we have to rebound.

Second, IT-clefts may be of *Type B*, i.e., *informative-presupposition cleft*, where the focus is restricted to the embedded clause. This type is especially frequent with clefted adverbials. One subtype, frequently found with manner or cause adverbials, is also called the *cohesive cleft* (strongly preferred in French and Portuguese, but almost non-existent in German and not so frequent in Italian or Spanish; cf. also Van den Steen 2005; Roggia 2008):

- (7) It is for this reason that [...] the Commission has decided to make 2004 the European Year of Education through Sport. (Dufter 2009: 103)

Finally, *Type C* IT-clefts can be called *all focus / all new* and are often found at the beginning of a discourse act or unit. They are almost non-existent in German. So-called *presentational clefts* or *HAVE-clefts* with this function, such as in example (8), are particularly common in French (cf. also Lambrecht 2001, ch. 4). They will thus be taken into account as a structural variant of *Type C* IT-clefts, even if this denomination might be slightly misleading (we are following Lambrecht 2001 here in using the English terms for non-English structures which are, however, functionally equivalent to them; cf. Lambrecht 2001: 468):

- (8) A: Qu'est-ce qui t' arrive? (B is crying)  
           WH                   2s.OBJ 3s.happen  
           'What's wrong with you?'  
       B: J' ai ma mère qui est malade.  
           1s 1s.have 1s.POSS mother REL 3s.be sick  
           'My mother is ill'

Dufter (2009: 108–109) explains that IT-clefts are chosen only rarely in German to translate the English occurrences in the EUROPARL-corpus because German has the ability to permit *in situ* constituent scope marking, topicalization (V-S-

inversion) and scrambling, which is heavily restricted or non-existent in French (Dufter 2009: 109–110; for similar tendencies in Spanish vs. French see Van den Steen 2005; for comparable findings on French vs. Italian see Roggia 2008). Modal particles are also used in German to signal presupposed content (Dufter 2009: 110–112). These alternative structures make possible the overall avoidance of Type B and C in German, which indicates different conventionalization tendencies for different functions of Cleft sentences in different languages. These observations lead to our Hypothesis III: We expect more clefts in French than in Italian (see also Roggia 2008 for this assumption) and still slightly more in Italian than in the German / Swiss German text messages, especially where Type B and Type C-*it*-clefts are concerned.

Let us now summarize our three hypotheses:

#### Hypothesis I:

*it*-clefts and their equivalents are expected to appear quite rarely in text messages, as they are typical instances of informal communication.

#### Hypothesis II:

- a. WH-clefts should be particularly frequent in text messages;
- b. WH-clefts are in general rather infrequent in the French, German and probably also Italian subcorpora.

#### Hypothesis III:

We expect more clefts in French than in Italian and more in Italian than in the German / Swiss German text messages, especially where Type B and Type C-*it*-clefts are concerned.

### 3 Database: The Swiss SMS corpus

Our study is based on the newly established corpus *SMS4science.ch*. In the fall / winter of 2009 and in the spring of 2011, 25,947 authentic text messages were collected with the support of Swisscom, the most important telecommunication provider in Switzerland, as part of a larger international project (cf. [www.sms4science.org](http://www.sms4science.org)). The corpus comprises about 500,000 words. In total, 2,784 participants sent their original text messages to a free phone number. Of course, we cannot verify the “authenticity” of the data (i.e., we cannot exclude that someone wrote a text message precisely for the study, thus not or less reflecting his or her normal usage), but we have to accept this general methodological problem for any kind of empiri-

cal study where speakers or writers know that their language products will be analyzed. The messages have subsequently been anonymized and are currently set up in a corpus that constitutes an online resource for scientific needs (cf. Dürscheid and Stark 2011 for further details on the corpus). Out of the 2,784 participants, 1,316 filled in an anonymous online questionnaire containing sociodemographic information about their sex, age, regional origin, mother tongues, education, reading and writing habits etc. (cf. [www.sms4science.ch](http://www.sms4science.ch) for overall information about the average sex, age etc. of the participants). Furthermore, 10,718 text messages are written in Swiss German dialects, 7,224 in non-dialectally marked German, 4,627 in French, 26 in a French dialect or Francoprovençal, 1,481 in (also dialectally marked) Italian, 1,085 in Romansh and 539 in English (the remaining 247 text messages contain Slavic languages, Scandinavian languages etc.).

For this study, we searched manually for Cleft constructions of all types in the Swiss German, non-dialectal German, French, Italian and English text messages.<sup>2</sup> There were no Cleft constructions found in the English text messages, which leaves us with the following four languages and varieties, respectively, constituting the base of our findings:

**Table 1:** Number of text messages / words per language / variety with Cleft constructions

	SMS	Words (approximately)
Swiss German	10,718	218,000
Non-dialectal German	7,224	140,000
French	4,627	94,000
Italian	1,481	30,500

The following examples<sup>3</sup> illustrate the kind of linguistic variation we find in the text messages from Switzerland:

- (9) oh schnuggiputz hallo, i miss you! Du hesches bald gschafft :- ) vernünftg bisch mitem zug gange, super! Cu soon :-\* . (Swiss German, English)  
'Oh Darling, hello, I miss you! You have almost made it :- ) It was reasonable that you took the train, super! See you soon :-\*'

<sup>2</sup> The manual research by reading each text message was effectuated by Alexandre Torea, a master's student at the University of Zurich, whom we thank for this valuable empirical work.

<sup>3</sup> The English translations of the text messages remain sometimes voluntarily non-idiomatic and very literal in order to make clear the original structure of the messages.

- (10) Na, hats bei euch auch geschneit? Ich kann heut abend so gegen 20.30 bei euch sein. Schick mir doch ne kurze Antwort, wenn dich meine SMS erreicht hat. CU Jörg (non-dialectal German, English)  
'So, did it snow at your place as well? I can be at your place at around 20.30. Please send me a short answer when you have got my text message. See you, Jörg'
- (11) Mé fiche de Geo son ché toi? t la se soir pr ke je viene les chercheR? (French)  
'Are my notes of geography at your place? Are you at home this evening so that I can come and get them?'
- (12) Hey scusami ho visto il tuo sms solo oggi. Bè è come l'hai detto te, la lingua nn è x niente facile. Pure io sono convinta che l'abbia dato a qualche persona (Italian, English)  
'Hi, sorry, I have only seen your text message today. Well, as you said, the language is not easy at all. And I am convinced as well that s/he has given it to some other person'
- (13) Joey angel, am still @mymother's place where we had dinner w my parents. Miss u, xxx sss (English)

As illustrated in these examples, text messages deviate from the written standard, especially with regard to register-specific spelling strategies such as the absence of apostrophes or the integration of apocopes, imitating the spoken language (cf. German *hats* instead of *hat's* for *hat es*, 'has it', or *heut* instead of *heute*, 'today', in [10]); "phonetic spelling" (cf. e.g. Anis 2007), particularly frequent in French, where the phoneme-grapheme-relation is very indirect and where silent consonant letters are very frequent and accordingly omitted in text messages (e.g. *mé* for *mes*, 'mine', *son* for *sont*, 'are', etc., in [11]); letter-and number-homophones (cf. Thurlow 2003), like *t* for homophonic *t'es*, informal French for *tu es*, 'you are', in [11]); consonantal skeletons (like *pr* for *pour*, 'for', in [11]); logographs like *x* for *per*, the mathematical operation, but also the preposition *per*, 'for', in Italian (cf. example [12]), and all kinds of abbreviations like *w* for *with* in (13). This enumeration is far from being exhaustive, and we will not go into the details of these specific spelling strategies (cf. for more information, see Dürscheid and Stark 2013), as they do not directly affect the number, distribution or formal set-up of the Cleft sentences we have found in our corpus. The empirical results of this study are presented in the next section.

## 4 Results

### 4.1 Overall results

The following table shows, in absolute numbers, all Cleft constructions and Cleft construction types found in the Italian, non-dialectal German, Swiss German and French text messages of our corpus:

Table 2: Overview of Cleft construction types in the corpus

	IT-clefts	WH-clefts	Reverse WH-clefts
Italian	5 (4+1 <i>have</i> )	–	–
Non-dialectal German	1	2	–
Swiss German	9	1	–
French	55 (38+17 <i>have</i> )	3	–

As displayed in Table 2, the absolute numbers are very low, with one exception: French, where we find 55 IT-clefts, among which 17 are formed with *avoir* (*Hello tu pourrai m excuser auprés de Chloé pq q un genou qui me fait vraiment mal* [hello 2s could 1s.OBJ apologise to Chloé because 1s.have ART.INDEF knee REL 1s.OBJ hurt much] 'Hello, could you excuse my absence to Chloé because I have a knee that is really aching...'). A typical example with a clearly contrastive focus on the clefted constituent is the following one from our SMS corpus:

- (14) **Cen'est pas ce jeudi que jem'absentepour voir les anciennes photos de Rafz**, mais jeudi prochain et Goran sera là. Bonne soirée! Patricia (French, IT-cleft)  
**It is not this Thursday that I will be off** to watch the old pictures of Rafz, but next Thursday, and Goran will be there. Good night! Patricia'
- (15) Liebe Florina, **was ich dir anbieten kann sind Karten** für die letzten beiden Vorstellungen am 5./6. Dezember, ginge das? PS [...] (German, WH-cleft)  
'Dear Florina, **what I can offer you are tickets** for the last two presentations on December 5<sup>th</sup> or 6<sup>th</sup>, would that be fine? PS [...]'

This German example is a clear case of a WH-cleft, a type that is rare in the corpus and even more so in the Romance subcorpora (see above, Table 2).

It is important to note that clefts are generally very infrequent in our corpus, and that they only appear in French in slightly over 1% of the respective text mes-



sages (see Table 3). This is perfectly in line, for example, with the findings of Roggia (2008:15), who compares French and Italian clefts in the corpus C-ORAL-ROM:

Table 3: Percentages of text messages with clefts

	IT-clefts	WH-clefts	Reverse WH-clefts
Italian	5/1,481 = 0.34%	–	–
Non-dialectal German	1/7,224 = 0.014%	2/7,224 = 0.028%	–
Swiss German	9/10,718 = 0.08%	1/10,718 = 0.01%	–
French	55/4,627 = 1.19%	3/4,627 = 0.07%	–

## 4.2 Functions of IT-clefts

If we have a closer look at the examples, it is sometimes difficult to decide which type of IT-clefts (as those constitute the majority of our occurrences) we are confronted with: Type A, *stressed focus cleft*; Type B, *informative-presupposition cleft*; or Type C, *all focus / all new* (see above, Section 2):

- (16) **No nn sn io ke ti pacco sempre, 6 tu ke arrivi sempre al momento sbagliato** © (Type A, multiple-focus?)  
 'It is not me who snubs you always, it is you who is always arriving at a bad moment'
- (17) **Sisch im fall gar nöd s telefon, wo kaputt isch, sondern d sim charte!**  
 Ha jetz jensti grät usprobiert und alli hend di glich fehlermäldig age!  
 (Type A)  
 'It happens to be not the phone which is broke but the SIM card! I have tried many different phones, and all indicated the same error!'
- (18) **Coucou! c'est pas aujourd'hui qu'on va voir ce film à l'abc? c'est à quelle heure?** (Type A)  
 'Hi! Isn't it today that we will go and watch that movie at abc? At which time?'

We consider these three examples as belonging to Type A, i.e., the clefted constituent is in narrow focus. Examples (16) and (17) are cases where this is particu-

larly evident, as the contrast is made explicit in the second coordinated sentence. Example (18) is more problematic as the contrast is not easy to reconstruct, thus it could contain an emphatic focus just to highlight the fact that the appointment in question will indeed take place that same day.

Example (19) belongs to Type B, i.e., a type of IT-cleft in which the focus is on the embedded sentence rather than the clefted constituent, which takes up something that has been said before by means of an anaphoric elements (here: *ça*, 'that').

- (19) **C'est pour ça que je ne prends JAMAIS les trains le dimanche soir...**  
 Sinon tu vas? Et ton Damian? [...] (Type B)  
 'It is for that reason that I never take the train on Sunday evenings...  
 Will you go there? And your Damian?'
- (20) Oups, moby isch ja no. Grad temporär vergässä, tztztz- ja, bi glaub chli zimli müed :- ) wohnig nöd okay, nun- **das isch nöd die gsi wo uf dich wartet**, ganz eifach :- ) gmüetliche abig u live bis i ca. 48std. Grueß vo frau fatigue (Type C)  
 'Oops, moby will be. I forgot that temporarily, tztztz – well, I think I am quite tired :- ) Apartment was not okay, well – **that was not the one which waited for you**, simply © cosy evening, and live in about 48 hours. Greetings from Mrs Tired'
- (21) Non, je suis entrain d'arriver, **c'est ma maman qui m'amène!** j'te dis quand j'arriverà toute, becs (Type C)  
 'No, I am about to arrive, **it is my mum who gives me a ride!** I will tell you when I am there, see you in a minute, bye'

Examples (20) and (21) are examples of Type C, i.e., the information in the Cleft sentence is all new (*das isch nöd die gsi wo uf dich wartet*, Swiss German, example [20]; *c'est ma maman qui m'amène*, French, example [21]). In example (20), *die* ('it') is in a rather sloppy anaphoric relation to *wohnig* ('apartment'), not referring back to the visited apartment, but to the one and only apartment that is potentially waiting for the author of the text message. Additionally, the negation of the fact that the visited one was the one "that waited" for the author of the text message renders the proposition in the whole sentence new. In (21), this is even more evident, as the mother of the author of the text message is not aforementioned, nor is the fact that s/he is being given a ride.

- (22) ciao bel...un'ora fa sono uscita dopo 3 ore di esame di chimica...ahahah...divertente...-” .. vabbé..dovrei averlo passato.cmq..sono mega distrutta psicologicamente :-/ **ho la testa che scoppia**, e domani ho fisica...=( [...]  
(Type C, presentational *have*-cleft)  
'Hi beautiful one...one hour ago, I left after 3 hours of chemistry exams...ahahaah...amusing...-=-well..I should have passed it.. However, I am completely destroyed psychologically :-/ **I have the head that is exploding**, and tomorrow, it will be physics...=( [...]'
- (23) Yo,bien débuté la semaine?**ya samuel qui propose de prendre l'auto** pour ce soir.Tu aimerais venir avec? (Type C, presentational *have*-cleft)  
'Yo, did you have a good start to the week? **There is Samuel who proposes to take the car** for tonight. Would you like to join us?'

As for their information structure, examples (22) and (23) are quite similar to the preceding examples (Type C, *all new*); however, they show a morphosyntactic peculiarity: instead of an expletive subject and a form of the copula verb *to be*, we find either the first person or an expletive and a form of *to have* (*ho la testa che scoppia* [1s.have ART.DEF head REL 3s.burst] 'I have the head that is exploding', Italian, example [22]; *ya Samuel qui propose...*[EXPL.LOC.HAVE Samuel REL 3s.propose] 'There is Samuel who proposes...', French, example [23]) as the predicate of the clefted part of the sentence. This is quite frequent in Romance languages (*presentational cleft*, see above, Section 2).

If we ask ourselves how frequently these subtypes of IT-clefts occur in our corpus and in the respective subcorpora, we find the following results:

Table 4: Overview of IT-cleft functions in the corpus

	Type A	Type B	Type C
Italian	2 (40%)	—	3 (60%)
Non-dialectal German	1 (100%)	—	—
Swiss German	5 (55.6%)	—	4 (44.4%)
French	25 (45.46%)	4 (7.27%)	9+17 <i>have</i> -clefts (47.27%)

Thus, only French exhibits the whole range of possible IT-cleft functions in our data, and it shows even more *all new* and *presentational IT-clefts* (Type C) than the other two types, such as Italian (which has, however, very few attested examples). Types B and C have not been observed in our non-dialectal German text

messages. Up to this point, our results confirm the overall tendencies observed and partially explained by Dufter (2009) for the EUOPARL-corpus (see above, Section 2). What is striking, given the fact that Swiss German is a diatopic variety of German, is that Type C is quite frequent in Swiss German text messages, a fact that – to our knowledge – has never been stated before.<sup>4</sup>

In what follows, we will look more closely at French Type C occurrences in our corpus.

Example (24) demonstrates the classification difficulties we occasionally encountered (and which might have slightly altered the quantitative results presented in Table 4), as text messages are often embedded into longer dialogues which we ignore. Thus, we do not know whether *Marion* in example (24), below, has been mentioned before or whether it is part of a list of alternatives, and whether Jose being picked up by somebody is already given knowledge. If this was the case, we would face a Type C-occurrence here, but if Marion is one of several potential people to pick up Jose, the example could also be a Type A-occurrence of a French IT-cleft.

- (24) Hello. **C est marion qui uient chercher jose** ce soir. (Type A or C?)  
'Hello. **It is Marion who will come and pick up Jose** this evening.'

Having a closer look at rather unambiguous occurrences of Type C in the French subcorpus, we observe certain lexicalization tendencies, which have also been mentioned by Dufter (2009). Thus, 5 out of 9 Type C-*c'est*-clefts show the manner adverbial *avec plaisir/avec joie* ('with pleasure / joy') as the clefted constituent:

- (25) Ah c gentil! :) **C avec plaisir ke je serai venu** ms je pourrai pa etre la...Ms fais signe pr le prochain coup! Bonne journée, Gui  
'Ah, that is nice! ☺ **It is with pleasure that I would have come**, but I will not be able to be there...But please inform me next time! Have a nice day, Gui'
- (26) oki..ben j pense qu i va trouver Ça biza aussi..lol euh ou i suis tj s partante:-) **c'est avec plaisir de repasser une soirée avec toi:** hi bisous chérie Jtm<sup>5</sup>  
'OK, well, I think that somebody will find that strange, too...lol euh yes, I am still taking part :-) **It is with pleasure that I will spend another evening with you** ;'9 Hi, kisses, darling, I love you'

<sup>4</sup> Elvira Glaser, personal communication.

<sup>5</sup> Here, the Cleft sentence is structurally incomplete (it would have been *c'est avec plaisir que je repasse une soirée avec toi*). The author of the text message most probably mixed up two differ-

- (27) Quand tu aura un moment **c'est avec grand plaisir que je t'invite** :) ca serait trop cool! Gros bisous  
'Once you find a moment, **it is with great pleasure that I will invite you**  
☺ that would be too cool! Big kisses'
- (28) On peut arriver dans l'après-midi, et **c'avec joie que je te donne un coup de main!**  
'We can arrive in the afternoon, and **it is with pleasure / joy that I will help you a bit**'
- (29) Coucou Lexi, **c'est avec joie, que je viendrai demain après-midi** voir votre princesse, vers 16h00, youpiiii! Je n'aurai pas le temps de faire un cake, mais ça n'est que partie remise, promis;-) [...] 'Hi there Lexi, **it is with pleasure / joy that I will come tomorrow afternoon** and see your princess, around 16h00, yippee! I will not have the time to make a cake, but that is only suspended, I promise;-) [...]'

The formulaic character of these clefts is easy to detect.

Example (30) is lexically similar, but with a different information structure:

- (30) Dommage pour moi **c'est pas l'envie qui manque** d'y manger une fois ... mais je ne suis pas là je l'ai dit à Nicky hier au boulot je vous souhaite une bonne soirée (Type A or C?)  
'That's a pity for me, **it is not the will that is missing** to eat there once... but I will not be there I told Nicky yesterday at work I wish you a nice evening'

Here, the clefted constituent (the subject of the unmarked corresponding sentence *l'envie d'y manger ne manque pas*) also expresses an emotional state of pleasure / joy / will, but the negation in the clefted part permits an alternative interpretation of this *c'est*-cleft as Type A ('it is not the will, but the occasion...').

The following two examples are even closer to lexicalization of the whole Cleft construction than examples (25) to (29) above:

---

ent syntactic constructions (the clefted one and *c'est un plaisir de repasser une soirée avec toi*). Informationally speaking, the resulting expression is nevertheless clearly Type C of our *it*-clefts and will accordingly be considered as such.

- (31) **C toi ki voi**.ben passe me prendr ver NNNN, je devrai etr pret.tte (Type A or C?)  
'**It's up to you**, well, come and pick me up around NNNN, I should be ready by then'
- (32) Bon... **C'est toi qui vois**.. J'espere que ça ira mais mm a 2h du mat tu sais ke tu px venir hin ;) Prends soin de toi, bisous JE T'AIME  
'Well...**It's up to you**...I hope that it will be possible, even at 2 o'clock in the morning you know that you can come there ;) Take care, kisses I love you'

Even if we could still imagine a literal reading of (31) in the sense that the addressee of this text message (*toi*, 'you') is the one to decide, in contrast to the sender or somebody else, this interpretation is very improbable in (32). In our opinion, *C'est toi qui vois*, which occurs twice in our corpus, is lexicalized, just as *it's up to you* (*to decide*), an original Cleft structure also found in English.

## 5 Discussion

As we have seen in Section 4, we find relatively few occurrences of Cleft constructions in our data. This is certainly partially explained by the fact that text messages are graphical, but rarely conceptually written; they are also different from dialogical and argumentative speech in that direct interaction and online observation of the interlocutor's performance are not possible, among other aspects. They are short, as their original name tells us (*Short message service, SMS*), and only quite rarely deploy an internal textual structure where clefts are needed as coherence-creating devices.

French prevails in that it exhibits 58 Cleft constructions altogether (see Tables 2 and 3), a number far outranking the findings for Swiss German (10), Italian (5) and non-dialectal German (3). This is a confirmation of the often stated preference French has for clefts (see e.g. Dufter 2008, where he discovers about three times higher percentages of clefts in his French corpora compared to the study of Collins 1991 on English), but the extreme differences in the use of Cleft sentences certainly require further comparative studies on clefts in computer-mediated communication.

As for the structural sub-types, we have found only very few *WH*-clefts in our subcorpora (and relatively more in non-dialectal German than in the other subcorpora, as expected); *IT*-clefts constitute by far the most important type.

We have also found a surprising partial confirmation of the general observation (cf. Miller 2006) according to which the frequency of (IT-) Cleft sentences diminishes in the languages of Europe from West to East. We have found the most clefts in the French subcorpus, followed by the Italian, followed by the Swiss German and subsequently the non-dialectal German data, and these results correspond *grosso modo* to Dufter's (2009) study using the EUROPARL-corpus. If one considers any map of Switzerland, one will state that French is the language spoken in the Western part (*Romandie*), close to the Swiss German area. Germany and Austria, both regions where Standard German is also spoken as a mother tongue, are situated to the (North-)East of Switzerland. Perhaps the fact that the position of Swiss German results in intensive contact with French provides an explanation for the different figures we have found in our Swiss German subcorpus compared to the non-dialectal German one (see Tables 2 and 3, above). In fact, there are relatively more Cleft sentences in Swiss German than in non-dialectal German, and we have found even Type C IT-clefts (example [20]), which could not easily be translated literally into non-dialectal or Standard German.

Furthermore, the expected presence of presentational *have*-clefts in Italian and French is confirmed by our data (Type C; see Table 4 and examples [22] and [23]).

Type B is attested only in French.

Among the Type C occurrences in French, a considerable number of them show clear lexicalization tendencies (*c'est avec plaisir / joie que, c'est toi qui vois*; cf. examples [25] to [29] and [31] and [32]).

Even if these results are generally in line with, for example, Dufter's findings (2009), we must admit that we encountered several classification problems due to missing context (cf. examples [24], [30] and [31], at least).

If we compare our results to our initial three hypotheses (see above, end of Section 2), the following conclusions can be drawn:

Hypothesis I is confirmed, i.e., IT-clefts and their Romance and Germanic structural equivalents do not appear very frequently in text messages, at least not in our corpus. Apparently, the informal nature is not as important when it comes to the appearance of certain linguistic features or stylistic variants as the *medium* in which informal communication is conducted – results for negation marking in French text messages and chats show similar tendencies, in that the expected features of informal phonic French do not appear to the same extent in these informal, but graphic communication forms (cf. Van Compernelle 2008; Stark 2012).

Hypothesis II is not confirmed, but language-specific restrictions formulated therein are confirmed. Thus, we find some WH-clefts in the German subcorpus, but to a much lesser extent than expected, and the German-Romance contrast is borne out (see Tables 2 and 3). Furthermore, 2 out of 3 non-dialectal German cleft

constructions are WH-clefts, 1 out of 10 Swiss German, and only 3 out of 58 French clefts are WH-clefts, with no attestation of a WH-cleft in Italian at all.

Hypothesis III is confirmed. We do find significantly more clefts in French than in Italian and more in Italian than in German and Swiss German text messages, especially as far as Type B and C-IT-clefts are concerned, which is in line with Dufter's (2009) findings. What is particularly striking here is the strong prevalence of French clefts in our data, which represent more than 11 times the Italian and German occurrences (Fr. 58, It. 5, G. 3 occurrences), and more than 10 times the Swiss German occurrences. While the French preference for clefts over other focalizing and coherence marking devices is a well-known fact, the extreme quantitative imbalance in the usage of clefts in our data is still striking.

We leave the search for a possible explanation to future research, but would like to conclude this discussion section with a brief general reflection on the variationist profile of text messages.

Robust quantitative results for certain linguistic features of (French) text messages that allow further-reaching conclusions about their variationist profile are not available in great magnitude at this point in time, and systematic empirical sociolinguistic and stylistic research is still almost absent for the new electronic communication forms in French. Yet, some pilot studies such as Van Compernelle (2008), Stark (2011, 2012) and Stark and Riedel (2013) indicate that the medial aspect of text messages' graphical nature plays a greater role than expected. Consequently, despite the fact that we do find typical features of informal French in text messages (negation particle drop, absence of past participle agreement, subject-verb disagreement etc.), we find it to a lesser extent than in informal corpora of phonic French. Our findings on clefts and their distribution point in a similar direction: the varieties used in text messages are most certainly different from speech, but not as informal as often assumed (see the fact that more IT-clefts are found than expected, in particular more Type C IT-clefts than Type A IT-clefts). Of course, there is also the factor of the reduced textual characteristics of text messages, which renders a high frequency of clefts improbable, as these are often considered to be coherence creating devices. Very often, there is a reduced necessity or possibility for exhaustive / contrastive foci against a presupposed background, and there is a greater tendency towards *all new*-information.

All in all, we consider text messages in particular and the new electronic communication forms in general as a fruitful empirical domain to learn more about language usage and medium dependent and/or register variation.

## 6 Conclusion

Our investigation into the types and distribution of clefts in the newly established Swiss reference corpus of text messages, more specifically into its French, Italian, Swiss German and non-dialectal German subcorpora, has shown that our data confirms the French preference for clefts in contrast to the other three languages and varieties, but that clefts are in general quite infrequent in text messages, perhaps due to their graphic and not precisely textual nature. It has revealed that the *all new* subtype C of IT-clefts (following Prince 1978) prevails in French and Italian, and that Swiss German text messages show surprisingly more clefts than the non-dialectal German text messages, a fact not noticed until now and which needs further investigation and explanation. Finally, this study has revealed that text messages are not just “transcribed informal speech”, and that their graphic nature and specific communicative profile leads to a different distribution of certain variety markers or indicators when compared with informal phonic corpora, especially for French. Further research is required in order to understand more about language use and probably also language change happening in computer-mediated communication.

## References

- <http://www.frantext.fr>  
<http://www.sms4science.ch>
- Anis, Jacques. 2007. Neography: Unconventional spelling in French SMS text messages. In Brenda Danet & Susan Herring (eds.), *The Multilingual Internet: Language, Culture and Communication Online*, 87–115. New York: Oxford University Press.
- Blanche-Benveniste, Claire. 2010. Les pseudo-clivées et l'effet deux points. In Marie-José Béguellin, Mathieu Avanzi & Gilles Corminbœuf (eds.), *La parataxe. Structures, marquages et exploitations discursives*, vol. 2, 185–217. Bern: Peter Lang.
- Büring, Daniel. 2006. Intonation und Informationsstruktur. In Hardarik Blühndorn, Eva Breindl & Ulrich Hermann Wassner (eds.), *Text – Verstehen: Grammatik und darüber hinaus*, 144–163. Berlin & New York: Mouton de Gruyter.
- Collins, Peter. 1991. *Cleft and Pseudo-cleft Constructions in English*. London: Routledge.
- Cresti, Emanuela & Massimo Moneglia. 2005. *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam & Philadelphia: John Benjamins.
- Dürscheid, Christa. 2002. E-Mail und SMS – ein Vergleich. In Arne Ziegler & Christa Dürscheid (eds.), *Kommunikationsform e-mail*, 93–114. Tübingen: Stauffenburg.
- Dürscheid, Christa & Elisabeth Stark. 2011. SMS4science: An international corpus-based texting project and the specific challenges for multilingual Switzerland. In Crispin Thurlow & Kristine Mroczek (eds.), *Digital Discourse: Language in the New Media*, 299–320. Oxford: Oxford University Press.
- Dürscheid, Christa & Elisabeth Stark. 2013. Anything goes? SMS, phonographisches Schreiben und Morphemkonstanz. In Martin Neef & Carmen Scherer (eds.), *Die Schnittstelle von Morphologie und geschriebener Sprache*, 189–209. Berlin & New York: Mouton de Gruyter.
- Dufter, Andreas. 2008. On explaining the rise of *c'est*-clefts in French. In Ulrich Detges & Richard Wältereit (eds.), *The Paradox of Grammatical Change: Perspectives from Romance*, 31–56. Amsterdam & Philadelphia: John Benjamins.
- Dufter, Andreas. 2009. Clefting and discourse organization: Comparing Germanic and Romance. In Andreas Dufter & Daniel Jacob (eds.), *Focus and Background in Romance Languages*, 83–121. Amsterdam & Philadelphia: John Benjamins.
- Engelbrechtsen, Steinar. 2012. Norwegische Satzspaltungen und ihre analogen bzw. nicht-analogen deutschen Entsprechungen: eine korpusbasierte Untersuchung. *Deutsche Sprache* 40 (2), 124–150.
- Ferrari, Angela, Luca Cignetti, Anna-Maria De Cesare, Letizia Lala, Magda Mandelli, Claudia Ricci & Enrico Roggia. 2008. *L'interfaccia lingua – testo. Natura e funzioni dell'articolazione informativa dell'enunciato*. Alessandria: Edizioni dell'Orso.
- Gast, Volker & Daniel Wiechmann. 2012. *W(h)*-Clefts im Deutschen und Englischen. Eine quantitative Untersuchung auf Grundlage des Europarl-Korpus. In Lutz Gunkel & Gisela Zifonun (eds.), *Jahrbuch des IDS 2011*, 333–362. Berlin & New York: Mouton de Gruyter.
- Gundel, Jeannette K. 2008. Contrastive perspectives on cleft sentences. In María de los Angeles, J. Lachlan Mackenzie & Elsa M. Álvarez González (eds.), *Languages and Cultures in Contrast and Comparison*, 869–887. Amsterdam & Philadelphia: John Benjamins.
- Jespersen, Otto. 1937. *Analytic Syntax*. London: Allan & Unwin.
- Koch, Peter & Wulf Oesterreicher. 2011 [1990]. *Gesprochene Sprache in der Romania. Französisch – Italienisch – Spanisch*. Berlin & New York: Mouton de Gruyter.
- Krötsch, Monique & Annette Sabban. 1990. “Bleu, je veux” – Remarques sur la focalisation en français. *Zeitschrift für romanische Philologie* 96 (1), 80–98.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
- Lambrecht, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39 (3), 463–516.
- Miller, Jim. 2006. Focus in the languages of Europe. In Giuliano Bernini & Marcia L. Schwartz (eds.), *Pragmatic Organization of Discourse in the Languages of Europe*, 121–214. Berlin & New York: Mouton de Gruyter.
- Prince, Ellen F. 1978. A comparison of *wh*-clefts and *it*-clefts in discourse. *Language* 54 (4), 883–906.
- Roggia, Carlo Enrico. 2008. Frasi scisse in italiano e in francese orale: Evidenze dal C-ORAL-ROM. *Cuadernos de Filología Italiana* 15, 9–29.
- Stark, Elisabeth. 2011. La morphosyntaxe dans les SMS suisses francophones: Le marquage de l'accord sujet – verbe conjugué. *Linguistik online* 48.
- Stark, Elisabeth. 2012. Negation marking in French text messages. *Linguisticae Investigationes* 35 (2), 341–366.
- Stark, Elisabeth & Isabelle Riedel. 2013. L'accord du participe passé dans les SMS francophones du corpus SMS suisse. *Romanistisches Jahrbuch* 63 (1), 116–138.
- Dürscheid, Christa & Elisabeth Stark. 2013. Anything goes? SMS, phonographisches Schreiben und Morphemkonstanz. In Martin Neef & Carmen Scherer (eds.), *Die Schnittstelle von Morphologie und geschriebener Sprache*, 189–209. Berlin & New York: Mouton de Gruyter.

- Thurlow, Crispin. 2003. Generation Txt? The sociolinguistics of young people's text-messaging. *Discourse Analysis Online* 1 (1). 1–27.
- Van Compernelle, Rémi A. 2008. Morphosyntactic and phonological constraints on negative particle variation in French-language chat discourse. *Language Variation and Change* 20 (2). 317–339.
- Van den Steen, Kathleen. 2005. Cleft constructions in French and Spanish. In Nicole Delbecque, Johan van der Auwera & Dirk Geeraerts (eds.), *Perspectives on Variation: Sociolinguistic, Historical, Comparative*, 275–290. Berlin & New York: Mouton de Gruyter.